ABSTRACT OF THE DISCLOSURE

A method for selecting textual content within a document. Text is selected using mechanisms of pattern recognition on the document's structure or content itself. A pattern recognition rule selects the desired text by identifying the start and/or end positions of the content in the document. The delineated contents is then said to be enclosed in an envelope. A series of envelopes may be used to identify the desired content. Successive envelopes are defined relative to a previous envelope. The contents of any envelope within a series, including the final envelope, may be extracted for use by other documents.